QA
76
M94
SI-Ref

M U D P I E no. 26

Museum and University Data, Program and Information Exchange

NATIONAL SCIENCE COMPUTER NETWORK

The National Science Foundation has announced (July, 1972) an expanded research program relative to the development of a national science computer network, which is described in announcement number NSF-72-16. The following paragraphs are taken from that announcement, since they are particularly pertinent to any future exchange of data, programs, and information among museums and universities, and should therefore be of interest to MUDPIE readers.

"BACKGROUND.--For several years the National Science Foundation has been sponsoring research generally applicable to the development of a National Science Computer Network. Such a network would provide its users with access to computing facilities, science information systems, and other computer-based resources on a regional and national basis. A National Science Computer Network would have profound implications for resource sharing in research and education. Facilities and resources of higher quality than might otherwise be the case could be available, and the duplication of some facilities would be eliminated. In particular, a national network would offer the possibility of integrating computing and science information services in ways to strengthen the Nation's research and education programs...

"...A comprehensive and interrelated set of project activities will be undertaken to permit exploration and evaluation of the many dimensions and facets of a nationwide network. Special attention will be given to collaborative efforts, independent of institutional affiliation, for development of innovative and effective approaches to resource sharing....The expanded program will provide for a trial National Science Computer Network utilizing existing computer communications technology and computer resources to support these network activities....

"...(The) network is envisioned as permitting institutions from every sector of our society to participate at minimum cost and without Federal subvention...A series of discussions and preliminary proposals is suggested which will permit the Foundation to gauge the community's sense of direction, interests, and requirements.

"USER FUNCTIONS:...(include)...design and use of large-scale information and data systems,...collaborative efforts on computer-based research methodology, including, for example, advanced computational techniques and hierarchical computing and informa-

tion services,...developing, testing and evaluating specialized computer-based instructional systems and materials,...innovative approaches to effective user integration of research and educational computing activities..."

It is apparent from this series of abstracted statements that NSF is working along the same lines as long since visualized as profitable by MUDPIE participants. I have, therefore, prepared a statement of what MUDPIE is all about, and have submitted it along with other pertinent materials to the Foundation, as a preliminary statement of interest. If the Foundation indicates a wish to explore further with us, we should be prepared to do so. Some of you have been associated with MUDPIE since its inception, and I know your interests, activities, and areas within which you would be willing to participate. Others among you have done little more than to ask to be put on the list for the newsletter, and I cannot say very much about your interests. Since we are now embarking on what might be the first steps toward implementation of a functioning network, I need to know what each of you might be willing to do in the way of participation. If you are already using time-shared computing, this might be a simple matter of indicating a willingness to switch to a mutual contractor so the network could immediately become a reality. If you are currently doing nothing, because your institution cannot afford a teletype connection for you, how much would you need to take this first step, if some help were forthcoming? If you are already well involved in time-sharing, are you interested in some kind of supervisory role in a planned network, to aid in development and growth (a MUDPIE Council?).

All of this should in no way interfere with the direct, individual participation of anyone in the proposed NSF program. I would like to know about other projects that will be suggested to NSF and submitted for support, and I will be glad to have information for future MUDPIES concerning such projects. Keep in mind that MUDPIE now has about 300 readers, among whom may be several who might be of value in any project.--JAP

MUSEUM AND UNIVERSITY DATA, PROGRAM AND INFORMATION EXCHANGE

The Museum and University Data, Program and Information Exchange had its origin in 1967, as a result of the growing interest at several institutions in the use of time-shared computers in museum research and in the potential for development of an inter-museum network for sharing information. The primary focus of attention has been on work in systematic and taxonomic biology, but much of the work done to date has had widespread application throughout biological research, and interest in the exchange (hereafter referred to as "MUDPIE") has been such that the original emphasis has been widened to include associated areas of investigation such as ecology.

Basically, MUDPIE represents a network, at present held together by its newsletter. The plan has been, since the inception of MUDPIE, to take advantage of commercial time-sharing computer facilities for direct exchange of ideas and uses of the computer as designed and implemented by the participants. There are at the present time many users of commercial facilities among museums, with fewer, perhaps, in the universities. The latter is due to the fact that universities have moved more strongly in the direction of in-house availability of computer laboratories for use by the professional staff in their research. Museum scientists, however, found commercial time-sharing a blessing, since it provided them with access to computers in a fashion usually not previously available to them. In many cases, museums have been exchanging programs and data files which are stored in local facilities, creating a considerable duplication and added expense that could easily be eliminated through the use of common availability to all museum users of a centralized memory storage capability, as would be provided by a commercial contractor with nationwide access to his central computer. This is, of course, presently possible, since many contractors offer whole or partial service to the nation, and on several occasions participants in MUDPIE have already taken advantage of this potential. The American Museum of Natural History and the United States National Museum had mutual access to each other's stored data and programs for a considerable period while both were contracting with the company CEIR, located in Washington, D.C., and the arrangement was discontinued only when the company revised its corporate structure, and both museums moved to new contractors. The National Museum and Michigan State University have had mutual storage access through the GE time-sharing system, which is national in scope, but rather expensive for most museum users. Demonstrations of the capacity and potential of time-sharing in museum work have been made in all parts of the country, using either the connection of the Smithsonian with Dialcom, a D.C. company, or with GE, through Michigan State University.

It has been a basic assumption for MUDPIE that the only material stored for full-time, mutual access in the network would be that for which use is highly repetitive. There are many collections of facts or organized sets of data that are constantly used by biologists in their day-to-day work or research, and most of these are amenable to computerization and on-line storage, given sufficient storage space. Among these are programs which permit standard calculations of biometric parameters and statistical correlations; taxonomic data matrices and programs for processing them; standard tables of biological data; answers to repetitive queries from the public; and many others. Progress in several of these areas has already been made by workers participating in MUDPIE, and functional programs, matrices, and answering services have been tested and are working. To take them up in turn, and in detail:

1. FUNCTIONAL PROGRAMS. Approximately 60 programs written for use by biologists in general and systematists in particular have been listed in various issues of MUDPIE. All such programs are available from the programmers, but at the present time must be sent through the mails as paper or magnetic tape or as a print-out. In addition, the Smithsonian has published a set of programs correlated with Simpson, Roe and Lewontin's book on quantitative biology. The set of programs (copy attached) has been distributed to all MUDPIE participants, and has also been added to the library facilities of several time-share companies, including one in Australia (used by the National Museum in Sydney). The Smithsonian has also published a program to permit analysis of biogeographical data in such a way as to distinguish biotically distinct areas, which, although originally designed to permit systematists to work with collections, has proven to be of use in pollution work, and has been adopted by environmental investigators. These are the kinds of programs, highly repetitive in their use, which should be centrally stored for use by biologists in all parts of the country, rather than individually stored on literally dozens of independent discs. During the time of mutual sharing by the American Museum and the National Museum, costs to each were reduced through the mutual access to programs prepared by the other, and a national network would correspondingly reduce costs for all those participating.

2. TAXONOMIC DATA MATRICES. The concept of storage of highly organized and coded taxonomic data has been fully developed by MUDPIE participants. The matrix, as currently designed, will permit storage of any and all kinds of data pertaining to a taxonomic group, regardless of size or variation of the group. It is organized in such a way that programs written either in BASIC or FORTRAN can access the data. It compresses an enormous amount of data into a comparatively small space, and permits rapid retrieval in many different forms. For example, one of the most repetitive activities of a systematist is the identification of unknown specimens both in the course of his own work and as a service to nonsystematic users of taxonomic data. The taxonomic matrix can be used as a random entry key to process information concerning the characters of unknowns, taking advantage of the computer's speed in surveying the entire range of knowledge concerning all taxa concerned. It is a swift, efficient, and highly accurate method (within the limits of the accuracy of the data as originally stored) for eliminating one of the most time consuming tasks of systematics. In addi-- tion, the same matrices are already being used for preparation of dichotomous keys which are printed by the computer; the printing of diagnoses or descriptions of taxa; preparation of contrasts between or comparison of two taxa, giving differentiating and similar characters; the preparation of "peek-a-boo" random entry field keys on punched cards; quizzing of students on their understanding of taxonomic groups; organization of groups showing

similar constellations of characters for phyletic and phylogenetic studies; and so on. The taxonomic matrix is amenable to practically every standard and some highly innovative techniques used by systematists, and is therefore both highly useful and repetitive in nature. We anticipate continued preparation and maintenance of matrices by specialists working with individually selected groups, since this is what is currently taking place. Thus, a matrix including all the genera of snakes from Central and South America is maintained in the Division of Reptiles of the National Museum, with constant revision and addition taking place as new information becomes available, and the matrix is currently functional as an identification tool. It is maintained under the contract of the Museum with a private contractor, but is available for use by workers anywhere in the United States right now, under special circumstances. Were there a national biological network it would still be maintained at no expense to anyone else, since it is a current research tool for the staff. It is expected, of course, that those with matrices in other areas would make them equally available at no cost to the National Museum, and this should be part of the thinking with regard to a national network.

Several papers and abstracts have been published concerning the taxonomic matrix, and additional material is in press. Copies of available materials were attached to the NSF material.

3. REPETITIVE QUERIES. Since the thinking with regard to MUDPIE has been primarily by museum biologists, it is natural that some work has been done in the direction of use of time-shared computers in the public exhibit area. Both the American Museum and the National Museum have prepared operational systems which have been actually tested in exhibit areas, and there can be no question but that it is completely feasible to make time-sharing a direct adjunct to and an educational extension of the standard museum exhibit. A manuscript concerning this use has been submitted to the "Museum Journal," and a copy of it was attached to the NSFmaterial.

This particular device has been of considerable interest outside the museum research community, since it is of fairly obvious use and significance to museums which are strictly exhibit oriented. Our original thinking was that each museum could take responsibility for a sector of knowledge, and each would make his file available to the other participant museums. This was demonstrated as viable by the two museums mentioned above, with the American Museum writing a file on mammals, and the National Museum doing one on reptiles and amphibians. The two files, stored with CEIR, could be accessed from either museum, and questions on mammals originating in the National Museum were easily answered from the American Museum file. It seems very likely that this system could be implemented in such a way that non-participant organizations, such as trail-side museums, would pay a small fee for each access to the files,

and thus reduce the overall costs for the active participants. There are several indications that enthusiasm would develop rapidly were this educational aspect of a network to be developed.

Since the beginning of MUDPIE we have offered instruction in the programming of time-shared computers, using the language BASIC. The course has usually been given in early summer, and has on several occasions been coordinated with the Systematics Seminars sponsored by the National Museum. About 15-25 people have taken the course each year. It has been run for two days, with the first day devoted to learning to program, and the second day spent in active contact with the computer, testing programs written by students. The courses have demonstrated that the material can be learned in a comparatively short time by a trained biologist, after which he is ready to participate in a network. Such instruction would be continued, of course, if the network were to be implemented.

One meeting of individuals interested in the potential of the Exchange has already been held, at the American Museum, in 1969. There were about 60-70 people present for the one-day session, which included demonstrations by time-share contractors of their equipment, additional demonstrations by scientists of the various on-line techniques in use at the time, talks by individuals who had developed ideas for time-sharing use, and round table discussions of the potentials for a network as well as continued exchanges and collaboration. A symposium to be presented at the First International Congress of Systematic and Evolutionary Biology, in Boulder, Colorado, in August, 1973, is devoted almost entirely to MUDPIE activities, and might well be considered the second meeting of participants.

AVAILABLE PUBLICATION

Sufficient copies of the publication " A New Approach in the Analysis of Biogeographic Data," Smithsonian Contributions to Zoology, no. 107, 1971, pp. 1-28, are available to supply the mailing list of MUDPIE, but many of you already have it, and I do not wish to duplicate extensively. If you want a copy, please drop a card to the author, James A. Peters, at the address below.

September, 1972
Division of Reptiles and Amphibians
National Museum of Natural History
Washington, DC 20560